Duplicative activation mechanisms of two trypanosome telomeric VSG genes with structurally simple 5' flanks

Duplicative activation mechanisms of two trypanosome telomeric VSG genes with structurally simple 5′ flanks

Keith R.Matthews+, Paul G.Shiels§, Sheila V.Graham, Catriona Cowan and J.David Barry*
Institute of Genetics and Wellcome Unit of Molecular Parasitology, Glasgow University, Church street, Glasgow G11 5JS, UK

Received October 23, 1990; Revised and Accepted November 26, 1990 EMBL accession nos X54731, X54732

ABSTRACT

In the mammalian bloodstream, African trypanosomes express variant surface glycoprotein (VSG) genes from a family of long and complex telomeric expression sites. VSG switching generally occurs by the duplication of different VSG genes into these sites by gene conversion involving a series of 70 base pair (70bp) repeats in the 5′ flank. In contrast, when VSG is first synthesised by trypanosomes in the tsetse fly at the metacyclic stage, a separate set of telomeric expression sites is activated. These latter telomeres appear not to act as recipients in gene conversion. We have found that the structure of two such expression sites is simple, with very short 70bp repeat regions and very little other sequence in common with bloodstream expression sites. However, the two telomeres readily act as donors in VSG gene conversion in the bloodstream and we show for one a consistent association of the conversion 5′ end point with the short 70bp repeat region. These findings help explain why a very predictable set of VSGs is expressed in the tsetse fly and have implications for VSG gene conversion mechanisms.

INTRODUCTION

The life cycle of the African trypanosome alternates between the blood-feeding tsetse fly and mammalian hosts (1). In mammals, the parasite population can survive for long periods exposed to the hosts' specific and non-specific immune responses because of antigenic variation, a process by which the parasites express sequentially a number of antigenically distinct variant surface glycoproteins (VSGs) (2). Each trypanosome has a repertoire of about 1000 silent, basic copy VSG genes, most of which lie in long tandem arrays within chromosomes (3). For expression, these genes must be, or become, located downstream of an active promoter in a VSG expression site (ES), of which there may be 20 or so in the trypanosome genome (4). All ES so far studied are telomeric (5) and contain a number of expression site associated genes (ESAGs) lying between the VSG gene and the promoter, which is tens of kilobases upstream (4,6,7,8,9). The mechanism by which only one ES is active at a time, thereby allowing expression of only one VSG, is unknown.

The events which result in a VSG gene becoming resident in an active ES can be categorized depending upon whether or not they involve gene duplication. Those which do not involve gene duplication include, rarely, reciprocal exchanges between active and inactive VSG gene loci downstream of an active promoter (10) and, more commonly, *in situ* activation of VSG genes in distinct ESs (11). Duplicative activation, in contrast, is believed to involve homologous gene conversion, with a basic copy being duplicated into an ES via homologous sequences which flank most VSG genes, resulting in the generation of an expression linked extra copy (ELC) of the VSG gene and loss of the formerly active VSG gene from the ES. The 3′ limit of conversion occurs within the relatively conserved 3′ end of the VSG gene or within the subtelomeric repeats which abut telomeric VSG genes. To the 5′ side, approximately 1−2 kb in front of VSG genes, is a series of (A-T)-rich 70bp repeats, which have the potential to form a variety of DNA secondary structures (12,13,14,15). For basic copy genes which reside within chromosomes, there are typically 1−10 copies of this repeat (2,16) while within VSG gene expression sites, many hundreds are usually present (15,17) producing a restriction site 'barren' region. This region marks the 5′ limit of many duplicative transposition events (14,17) and it has been suggested that, by providing extensive homology, the number of repeats adjacent to chromosome internal genes may contribute directly to the frequency of gene conversion (18,19).

An exception to telomeric VSG gene loci containing 5′ barren regions is those used in the trypanosomes' metacyclic stage (20,21,22,23), which develops in the tsetse fly salivary gland and dons the VSG coat as a preadaptation to life in the mammal (24). The metacyclic VSG (M-VSG) repertoire is far more limited than the more complex system used in the bloodstream (25,26,27), is very predictable in composition (25,28) and is activated by a mechanism which is distinct from the bloodstream system (27). M-VSG genes also appear to have a rather

* To whom correspondence should be addressed

+ Present addresses: +Department of Internal Medicine, Yale MacArthur Center for Molecular Parasitology, Yale University School of Medicine, New Haven, CN, USA and §Division of Molecular Biology, The Netherlands Cancer Institute (Antoni van Leeuwenhoek Huis) Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

characteristic location on the telomeres of the parasites' largest chromosomes (21,22,23). We have found previously that there appears to be a stage-specific activation mechanism for two M-VSG genes. In metacyclic-derived cloned populations the genes are activated *in situ* (29), while they are activated exclusively by duplication in the bloodstream (22,29). Here we have examined the structure of the telomeres containing these metacyclic expression loci, with particular emphasis on how they are able to undergo gene conversion in the bloodstream. Our results have implications both for the distinctive features of the M-VSG system and for the mechanisms of VSG gene expression site conversion.

MATERIALS AND METHODS

Trypanosomes

The trypanosomes used in these studies, a virulent cloned line of *Trypanosoma brucei* EATRO 795 which retains tsetse fly transmissibility, have been described previously (29,30). Trypanosome transmission through tsetse flies, maintenance in mammalian hosts, cell cloning and DNA preparation were carried out by standard procedures (30,31,32). Trypanosomes expressing the M-VSG gene ILTat 1.22 via an ELC were isolated by re-cloning from the the first patent parasitaemia in mouse infections initiated by metacyclic trypanosomes or from infection initiated by bloodstream trypanosome injection. These 1.22 ELC expressors were 1.22a, 1.22b, 1.22c, 1.22d, 1.22e, 1.22f, 1.22g, 1.22h and 1.22j'. 1.22i was a metacyclic-derived clone which expressed the 1.22 VSG gene *in situ* (29).

Library construction, DNA clones, Southern hybridization and DNA sequencing

Clones derived from the basic copy gene locus for the 1.22 gene, were derived by the insertion of trypanosome genomic DNA, partially digested with Sau3A, into lambda EMBL4 (Promega Biotech) (33). The clone lambda MT1.22B was selected by using the 1.22-specific cDNA clone pTcV7.1-14 as a probe (22).

Clones derived from the two basic copy gene loci for the M-VSG ILTat 1.61 were isolated by the cloning of specific genomic restriction fragments into pUC19 for pMT1.61-1,4 and 6, and pBluescript for pMT1.61-5. Plasmids pMT1.61-X5, pMT1.61-BI and pMT1.61-X6 were derived as subclones of a lambda clone, lambda MT1.61, which was generated by the insertion of fully EcoRI digested trypanosome genomic DNA into Lambda Dash (Stratagene). All clones were isolated by a very limited form of the chromosome walk procedure, with the first clone, pMT1.61-1, being selected using pTcV7.15-21, a 1.61-specific cDNA clone (22).

The plasmid clone pMG7.1-1 contains part of the 1.22 gene and upstream flank (22). Probes for 70bp repeat sequence and expression site associated genes were prepared from the VSG gene 221 expression site-derived plasmids pTg221-2, pTg221-8 and pTg221-9 (5,34). The plasmid clone pTcTIM9 contains a triosephosphate isomerase cDNA sequence (35).

Southern hybridization was carried out using Hybond-N (Amersham International plc) or Nytran membrane (Schleicher & Schuell) according to the manufacturers' instructions. Radiolabelled probes were prepared by the hexanucleotide random priming method (36). DNA sequencing was performed with Sequenase according to the manufacturer's instructions (USB).

RESULTS

Structure of the metacyclic VSG gene loci

We cloned approximately 17kb of each of the basic copy loci for the genes encoding the M-VSGs ILTat 1.22 and 1.61. These basic copy loci are used as expression sites in metacyclic-derived cells (29). Maps of the cloned regions are shown in Figure 1. For both loci we verified that the clones are in full agreement with their genomic equivalents by mapping and comparing digests; there was no evidence of rearrangement or deletion (e.g. Figure 2). A notable feature of these M-VSG gene expression loci was a lack of repetition in the genome (stringency: 2×SSC, 55°C), which is very different from bloodstream VSG gene expression loci. Figure 2, panel A shows this lack of repetition for the 1.22 gene locus. Probe 1 in Figure 1, which contains the 1.22 VSG gene, hybridized to a 3.8kb band in both the lambda clone and the trypanosome genome (Figure 2 panel A, lanes 1 and 2). The additional band, of about 23kb in lane 2, is an ELC of 1.22, present in this particular trypanosome clone (1.22j'). Hybridization with probes 3 (lanes 3 and 4) and 4 (lanes 5 and 6) also demonstrated both the non-repetitive nature of the expression site sequence in the trypanosome genome and the fidelity of the cloned region.

In EATRO 795 there are 2 basic copies for the ILTat 1.61 gene (22,29). Extensive Southern analysis of these loci, probing with both the cDNA for the gene and all the genomic clones, revealed no detectable difference between the two loci at the level of restriction endonuclease mapping. As with the 1.22 locus, the majority of the cloned region detected just single bands when hybridized to genomic Southern blots at high stringency (0.1×SSC, 65°C) (Figure 3), although at lower stringency (2×SSC, 65°C) cross hybridization with a few other bands was detected (data not shown). Only one region was very repetitive at this lower stringency (Figure 3 panel B), but apparently unique at high stringency. Genomic maps for the 1.61 loci and derived plasmid clones were in full agreement.

Repetitive regions on the metacyclic expression loci

Although most of the sequence within the two loci appeared to be at low copy number in the genome, we detected two reiterated regions in each locus. Firstly, approximately 1.5kb upstream of both M-VSG genes there is a highly reiterated sequence (see Figure 2 panel B, probe 2 and Figure 3 panel A, probe 8). We found, by hybridization, that this was 70bp repeat sequence, as already indicated in a preliminary analysis (22). Figure 4 panel A shows that a 70bp repeat probe hybridises specifically to a 420bp PstI/PvuII fragment of lambda MT1.22B. Similarly, in the 1.61-derived clones, hybridization to 70bp repeat sequence was detected over a 900bp XbaI/HindIII fragment (Figure 4, panel B). We have carefully compared sizes of the cloned regions and the corresponding genomic fragments because of potential deletion or rearrangement of such repetitive sequences. We detected no discrepancy to a resolution of c. 50bp for the 1.22 locus and c. 100bp for the 1.61 loci (data not shown).

We sequenced the 70bp repeat-hybridizing regions from the two loci (Figure 5); for each there are less than 2 complete consensus 70bp repeat units (16). In the 1.22 locus, extensive similarity with the consensus breaks down approximately halfway into the second repeat, while in the 1.61 locus there are nearly 2 full repeats. Both sequences show a high degree of identity with the consensus, although in the 1.61 repeats there is disruption of the (T Purine Purine) triplets and of the second TGTTG box, which is a conserved feature of 70bp repeats (16). The brevity

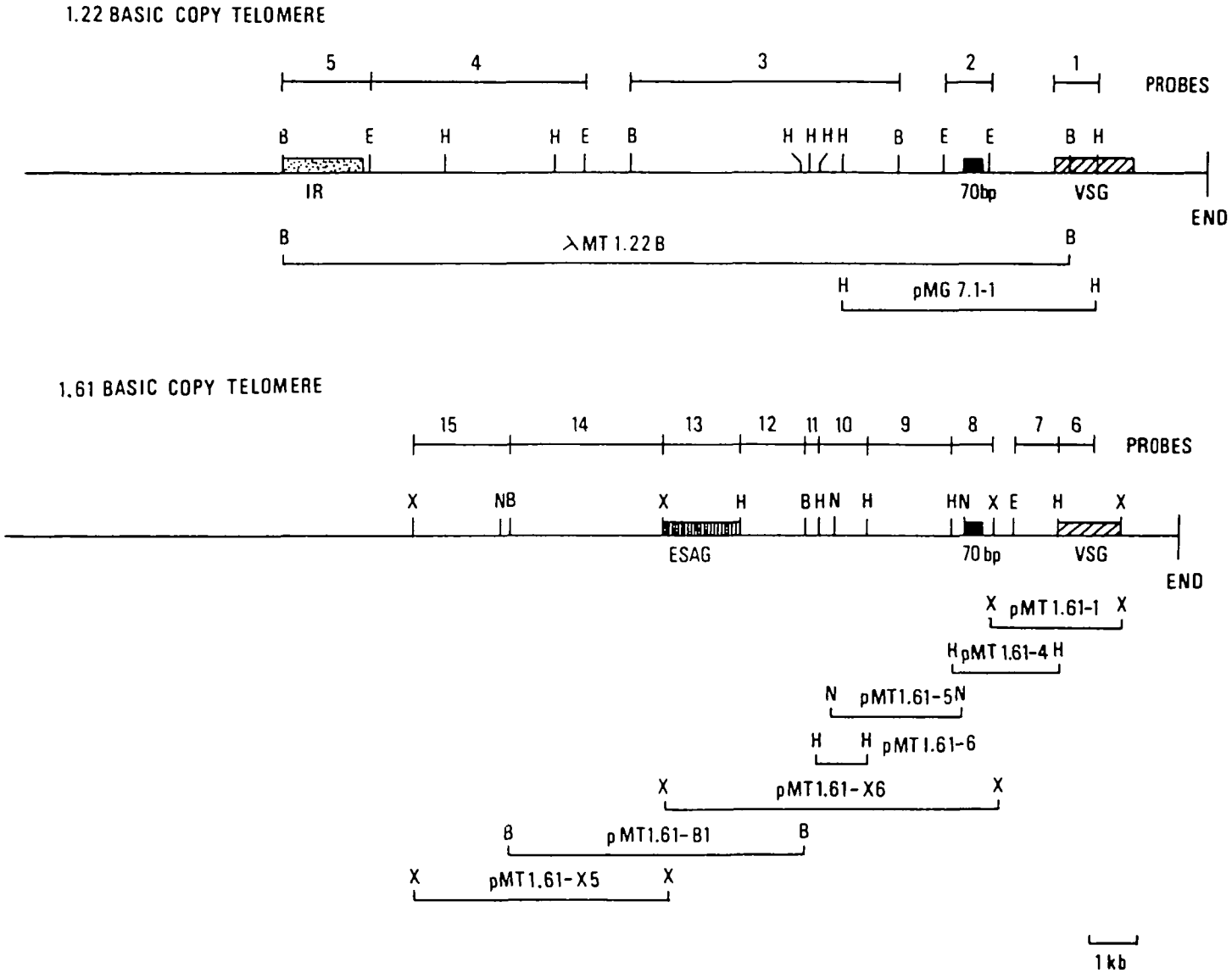

Figure 1. Restriction maps of the BC telomeres containing the M-VSG genes encoding ILTats 1.22 and 1.61, and the genomic DNA clones derived from them. Only one of the two telomeres harbouring the 1.61 genes is shown, their restriction maps being identical except for the distance between the VSG gene and chromosome termini. The abbreviations used for restriction endonucleases are:- B, BamHI; C, ClaI; E, EcoRI; H, HindIII; X, XbaI; N, NaeI. Black boxes mark the positions of the 70bp repeat unit regions (70bp) (see Fig.4); stripped box; expression site associated gene 1-related sequence (ESAG) (see Fig.6); stippled box; partial Ingi retroposon sequence (IR); hatched box; VSG gene region (VSG); END, telomere. The regions from which probes 1–15 were derived are shown.

of the repeat arrays in these two M-VSG loci is novel with respect to bloodstream VSG gene expression sites, where they occupy many kilobases (13,15,17,18). On other M-VSG gene loci, no repeats whatsoever have been detected previously (21).

We have identified two other repetitive elements. In the 1.22 locus, at the limit of our available clones, is a sequence which detects many bands in the trypanosome genome, even at high stringency (see Figure 2 panel C probe 5). Although a full analysis of this area has been precluded by our consistent inability to walk beyond lambda MT1.22B, preliminary sequence analysis (O. Shonekan, unpublished results) has revealed extensive similarity with the 5′ end of the trypanosome retroposon-like sequence, Ingi (37,38).

In the 1.61-derived clones, about 7kb upstream of the VSG gene, there is an expression site associated gene 1 (ESAG 1)-related sequence (Figure 6, panel II). ESAG 1, which is present in other bloodstream and M-VSG gene expression loci, usually detects approximately 17–25 bands in genomic DNA at high stringency (6,8,39). Hybridization at high stringency, however, using the ESAG 1-related sequence from the 1.61 clone, detected only a single band (Figure 3 panel A, probe 13), although at long exposures c. 5 very weak bands were also detected (data not shown). This indicates that the ESAG 1-related sequence might be quite diverged from previously characterized ESAG 1 sequences; at lower stringency, the probe detects other fragments (Figure 3 panel B, probe 13). We have found no evidence of ESAG 1 in the 1.22 locus and in neither 1.22 nor 1.61 loci have we detected, even at low stringency, the second and third ESAGs found in bloodstream expression sites (Figure 6). In both loci we have searched for all ESAGs so far identified (9) and have detected no hybridization (data not shown). Furthermore, the lack of genomic repetition of sequences from both expression sites, even at moderate stringency (65°C, 2×SSC), makes it unlikely that any other ESAG-like sequences are present.

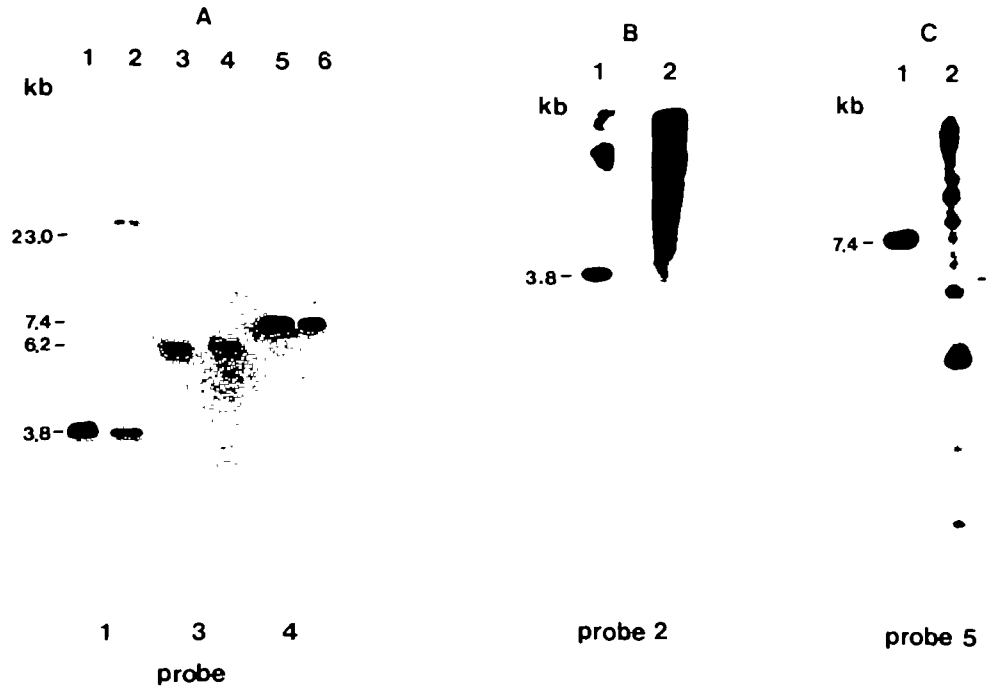

Figure 2. Analysis of 1.22 genomic clones. *Panel A*: Southern blot analysis to confirm the genomic authenticity of the lambda MT1.22B cloned sequence. Alternate lanes show BamHI digested lambda MT1.22B (lanes 1,3,5) and genomic DNA (lanes 2,4,6) probed with a series of fragments from lambda MT1.22B. Probes were: the 5′ fragment from pTcV7.1-14 (probe 1 in Fig. 1; lanes 1,2); a 6.05kb BamHI fragment from lambda MT1.22B (probe 3; lanes 3,4); a 4.7kb EcoRI fragment from lambda MT1.22B (probe 4; lanes 5,6). *Panel B*: A repetitive region in the lambda clone. BamHI digested lambda MT1.22B (lane 1) and 1.22 genomic DNA (lane 2), probed with probe 2. *Panel C*: A second repetitive region. Samples as in panel B, probed with probe 5.
All washes were to 0.1×SSC, 65°C.

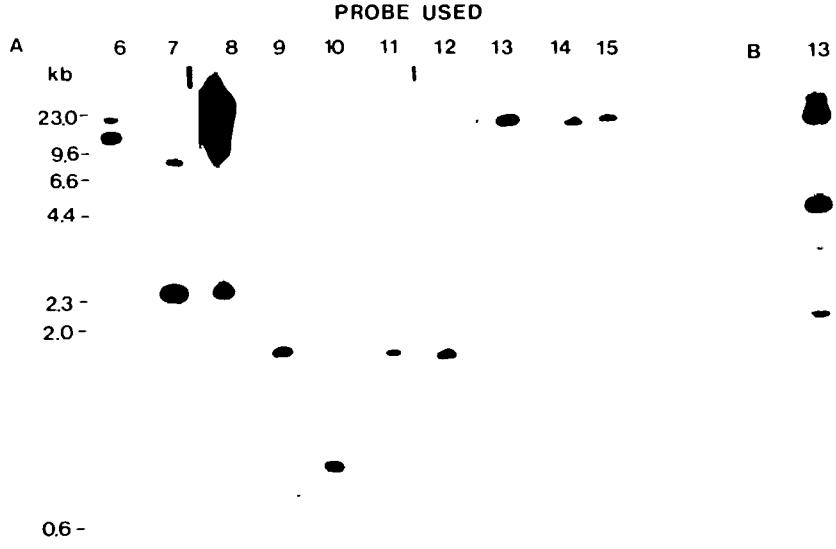

Figure 3. Southern blot analysis to demonstrate the absence of extensive genomic repetition over the regions cloned from the 1.61 gene loci. All lanes are HindIII digested 1.22 genomic DNA, hybridized with the probes illustrated (shown in Figure 1). All washes were to 0.1×SSC, 65°C. Panel B shows probe 13, which is repetitive at reduced stringency only, hybridized to HindIII digested 1.22 genomic DNA. This panel was washed to 2×SSC, 65°C.

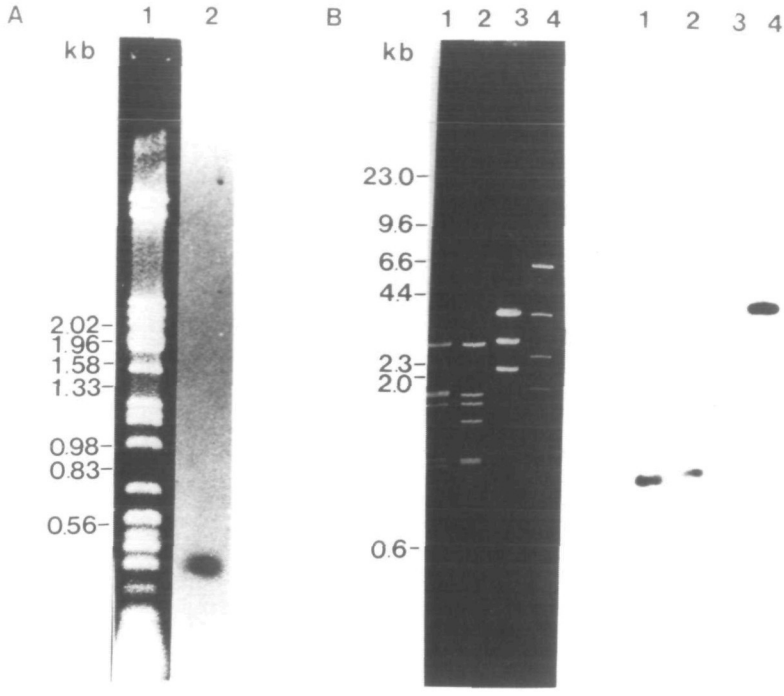

Figure 4. 70 bp repeats are present on both the 1.22 and 1.61 gene loci. *Panel A*: The 1.22 locus. On the left (lane 1), ethidium bromide stained gel fractionation of lambda MT1.22B digested with PstI and PvuII. On right (lane 2), blot of same gel probed with a 70bp repeat probe derived by PstI + TaqI digestion of pTG221-2 (34). *Panel B*: The 1.61 locus. On the left, ethidium bromide stained gel separation of pMT1.61X6 double digested with o′−6 HindIII + XbaI (lane 1) and with BamHI + HindIII (lane 2), pMT1.61X5 digested with BamHI + XbaI (lane 3) and pTg221-8 digested with PvuII (lane 4). On the right, a blot of the gel was hybridized with the same probe as used in panel A. Blots were washed to a stringency of 2×SSC, 50°C (for panel A) or 55°C (for panel B).

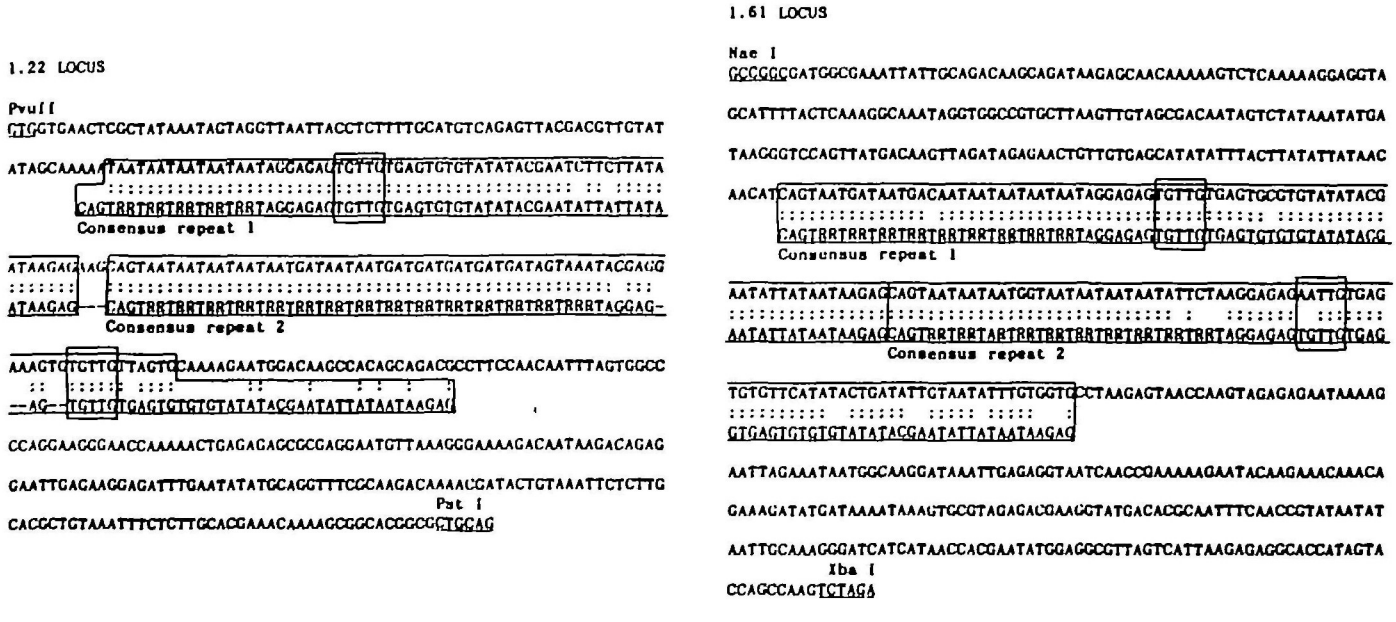

Figure 5. Sequence of the 70bp repeat-hybridizing regions on the 1.22 and 1.61 gene loci. For each, a colon indicates agreement with the consensus of Aline *et al.* (16). R = Purine nucleotide. Gaps have been introduced to maximize alignment between the sequence and the consensus. Repeat units are boxed. The conserved TGTTG motifs are also boxed.

Duplicative transposition of the ILTat 1.22 gene

We have previously found that the 1.22 and 1.61 M-VSG genes are activated without duplication in clonal populations derived from metacyclic trypanosomes (29). With activation in the bloodstream, however, an additional copy of both genes was generated, presumably involving the conversion of a bloodstream VSG expression locus. The relatively non-repetitive nature of these two loci prompted us to find out if the remarkably short 70bp repeat arrays on the M-VSG gene loci were sufficient to

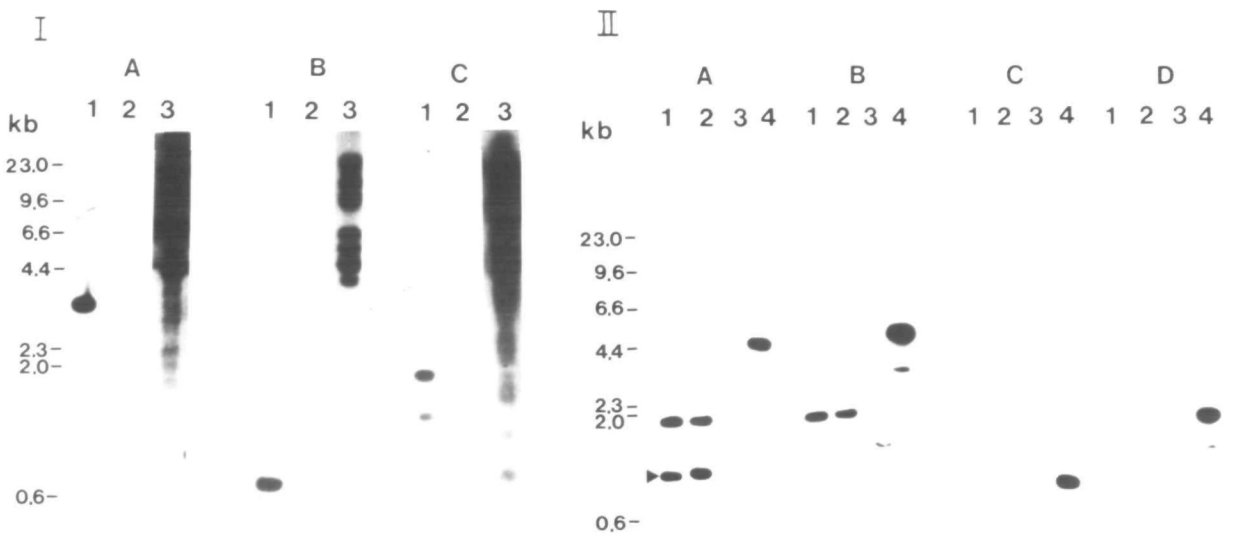

Figure 6. Southern blot analysis of ESAG homology to lambda genomic clones. *Panel I.* The 1.22 clones. *IA*: Low stringency analysis for ESAG 1. Lanes are: lane 1, control plasmid DNA: pTg221-8×PvuII; lane 2, lambda MT1.22B× EcoRI; lane 3, 1.22 genomic DNA×EcoRI. These were probed with a PvuII fragment derived from pTg 221-8 containing ESAG 1 and washed to (3×SSC, 42°C; Panel A). *IB*: High stringency analysis for ESAG 1. Same as IB, but washed to 0.5×SSC, 65°C. *IC*: Analysis for other ESAGs. Lanes are: lane 1, control plasmid DNA, pTg 221-9×HindIII, lanes 2 and 3 are as for IA and IB. These were hybridized with a HindIII fragment from pTg221-9 containing sequences representative of 2 other ESAGs. Hybridization and subsequent washing was at 3×SSC, 42°C. *Panel II.* The 1.61 clones. *IIA*: Low stringency analysis for ESAG 1. Lanes are: lane 1, pMT1.61-X6×HindIII + Xba I; lane 2, pMT1.61-X6× HindIII + BamHI; lane 3, pMT1.61-X5×HindIII + BamHI; lane 4, control plasmid pTg221-8×PvuII. These were probed with the same probe as in IA, then washed to 2×SSC, 50°C. The ESAG 1 probe cross reacts with the 70bp repeat containing region on the 1.61 loci (arrowed). *IIB*: High stringency analysis for ESAG 1. Same as IIA, but washed to 0.1×SSC, 65°C. *IIC*: Low stringency analysis for a second ESAG. Blot is duplicate of IIA, probed with a PvuII fragment of pTg 221-8 corresponding to probe A of Kooter *et al.* (6), then washed to 2×SSC, 50°C. *IID*: Low stringency analysis for a third ESAG. Same as IIA, probed with a PvuII fragment of pTg 221-8 corresponding to probe B of Kooter *et al.* (6), then washed to 2×SSC, 50°C

act in conversion events. We have examined this in detail for only the ILTat 1.22 gene because we have a large number of ELC expressors for this gene.

In all the bloodstream 1.22 expressors the 5′ portion of the 1.22-specific cDNA probe, pTcV7.1-14 (probe 1, Figure 1), detects, in addition to a fragment running from within the VSG gene to 5.3kb upstream, an extra band of variable size, representing an ELC (Figure 7). Exceptions are the control metacyclic-derived clone 1.22i (lane 7), which expressed the 1.22 gene *in situ* (29) and 1.22e (lane 5), which shows apparently two ELCs. We found no evidence of an ELC in the 1.22i population even at very long exposures of the autoradiogram shown in Figure 7; the small amount of DNA available for this lane resulted from the necessity to harvest metacyclic-derived trypanosomes as short a time as possible after leaving the tsetse fly. The two ELCs apparent in 1.22e may have arisen if the population originated from two independent trypanosomes, although it is possible that two separate conversion events generated the two ELCs, one or both of which may be transcriptionally active. In 1.22b (lane 2) there is an apparent smear above the distinct ELC derived band and this has been observed consistently. This might represent multiple ELCs or intrapopulation barren region size fluctuations at a single expression site. The multiple bands in 1.22e and 1.22b are unlikely to be due to incomplete DNA cleavage; rehybridization of the blot with a triosephosphate isomerase cDNA probe revealed that all DNAs were fully digested (data not shown). Thus, these experiments demonstrated that all of the

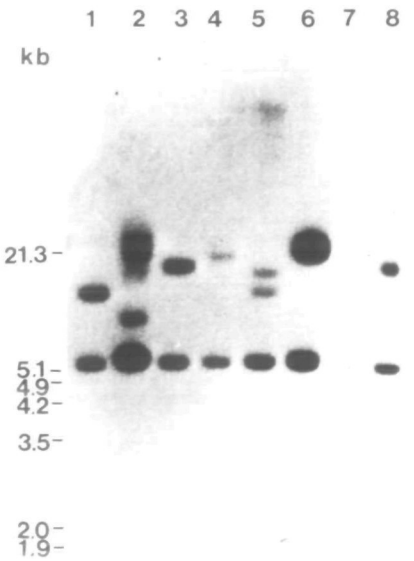

Figure 7. ELCs for the 1.22 gene are present in parasite clones initiated with the bloodstream stage. Lanes 1–8 are, respectively, 1.22a, 1.22b, 1.22c,1.22d, 1.22e, 1.22f, 1.22i, and 1.22j′ digested with HindIII and hybridized with the 1.22 specific cDNA pTcV7.1-14 (22). The blot was washed to 0.1×SSC, 65°C.

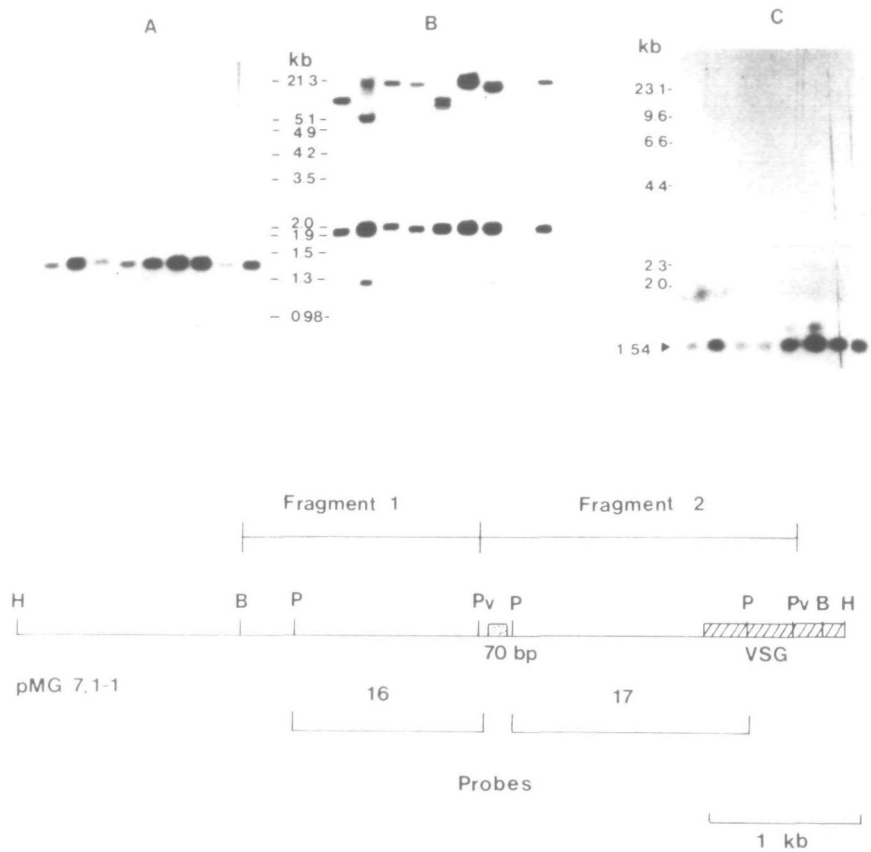

Figure 8. The 70bp repeat-containing region always delimits the duplicated region. Panel A shows probe 16 hybridized to the various 1.22 expressors used in Figure 7. Lanes 1–9 are, respectively, 1.22a, 1.22b, 1.22c, 1.22d, 1.22e, 1.22f, 1.22g, 1.22i, and 1.22j' digested with BamHI and PvuII. Panel B is the same blot stripped of its original probe and rehybridized with probe 17. Panel C is probe 17 hybridized with the same DNAs (excepting 1.22g; insufficient was available for this blot) digested with PstI. Lanes are, respectively, 1.22a, 1.22b, 1.22c, 1.22d, 1.22e, 1.22f, 1.22i, and 1.22j'. All blots were washed to 0.1×SSC, 65°C. Below is shown the origin of the probes used and the fragments detected, as described in the text.

populations, excluding the control *in situ* expressor, contained at least one ELC for the 1.22 gene.

We then examined whether the 70bp repeat-containing region on the 1.22 locus was associated with the 5′ limit of gene conversion by mapping round that region in the telomeres containing the basic copy and ELC genes. In Panel A of Figure 8 probe 16 detected only a single band in all DNA samples, which were doubly digested with BamHI and PvuII. This revealed that in none of the populations does the duplication limit lie within the region of the basic copy locus immediately upstream of the 70bp repeats (fragment 1, Figure 8). It also indicated that in all cases the DNAs were digested to completion. Panel B is the same blot stripped of the original probe and rehybridized with probe 17. In this case, all populations, with the exception of 1.22e (lane 5), 1.22b (lane 2) and 1.22i (lane 8), displayed two bands: the basic copy locus (2.0kb) and the ELC (variable size). As before, 1.22i has only its basic copy while 1.22e and 1.22b have additional bands. Therefore the duplication limit falls either within the 70bp repeat fragment or downstream (fragment 2, Figure 8).

Panel C of Figure 8 shows the DNAs digested with PstI and hybridized with probe 17. In this case, for each DNA, just a single band was detected, excluding the presence of a cross reacting band at about 1.6kb. This band, and the 1.3kb band seen in panel B, is due to cross-reaction with a region outwith the 1.22 expression locus which is present in both expressors and non expressors of the 1.22 gene (data not shown). Thus, panel C represents hybridization of the probe over a region of the 1.22 gene locus where the restriction maps for the basic copy and the ELC are the same. From this it is possible to conclude, for all ELC expressors examined, that the limit of the 1.22 gene duplicated segment lies upstream of the area equivalent to probe 17 (Panel C) and downstream of the area represented within probe 16 (Panel A); that is, the 420bp PstI/PvuII fragment containing the 70bp repeats. Finally, we mapped the ELCs for five of the populations analysed above to determine whether one or several expression loci could act as acceptor sites for the gene. We identified three distinct 1.22 ELC-containing loci (data not shown).

DISCUSSION

The expression loci of bloodstream VSG genes seem to have a common architecture (6,8,9). There is the telomeric VSG gene, an array of 70bp repeats extending over several kilobases and a number of expression site associated genes coordinately transcribed with the VSG gene. These common features, through recombination events, are believed to permit the frequent genomic

rearrangements which result in the trypanosome's ability to undergo antigenic variation. Here we have described two expression loci which lack extensive homology with previously characterized VSG expression telomeres, but which are fully able to participate in duplicative transposition events.

Structure of the metacyclic expression loci

The two M-VSG gene expression sites have a relatively simple structure. They lack the extensive genomic reiteration seen for bloodstream expression loci, and instead have only a strikingly short 70bp repeat array in common. The extent of repeats in both loci is the shortest that has been reported for telomeric VSG genes. Only in the case of other M-VSG gene loci has something similar been found: two were identified with no repeats whatsoever (21). In that case it was suggested that the lack of these sequences resulted in M-VSG genes being 'trapped' in their metacyclic-specific expression locus, so that only activation *in situ* would occur. However, although we have also found that there is activation *in situ* when our two M-VSG genes are expressed in metacyclic-derived organisms, this is not due to their lacking the signals for transposition; for example, the 1.22 locus does have a 70bp repeat which appears to be fully competent to act in conversion events.

The ability of the genes for ILTats 1.22 and 1.61 to undergo recombination events is significant because of the predictability of the M-VSG repertoire. The 1.22 M-VAT (Metacyclic Variable Antigen Type) is the most stable in the repertoire, being ubiquitous in trypanosome stocks isolated over a 17 year period in the field and also remaining stable during repeated fly transmission in the laboratory (28). Its basic copy gene locus is also relatively stable; in a number of trypanosome stocks we have detected differences only in the copy number of the entire telomere, rather than within the telomere (unpublished observations). This stability in the trypanosome genome is in contrast to the fluidity of the bloodstream VSG gene repertoire, and may be a consequence of the relatively non-repetitive nature of M-VSG expression sites. By lacking a long barren region and, at least for the 1.22 locus, having an absence of sequence in common with bloodstream expression sites, it may be that the M-VSG gene loci are relatively insusceptible to invasion by other telomeres. This recombinational isolation might assist the trypanosome in maintaining an M-VSG system functionally distinct from the more complex bloodstream system (29,39).

The organisation of the 1.22 gene environment may be at one extreme of M-VSG gene expression loci. The 1.61 gene loci were found to contain an expression site associated gene 1-related sequence, although this appeared relatively diverged from previously characterized ESAG 1 sequences at the level of hybridization. The presence of this area in common with bloodstream VSG gene expression sites might contribute to instability of the 1.61 loci in the metacyclic repertoire. Although apparently quite stable in field isolates, this M-VAT was lost from the metacyclic repertoire after only three rounds of cyclical transmission in the laboratory and did not subsequently reappear (28). The M-VAT 4 gene locus previously analysed appears to have a similar anatomy to that of ILTat 1.61 and also has been reported to be susceptible to deletion from the trypanosome genome (40). It is apparent, therefore, that M-VSG gene loci might show a range of relatedness with bloodstream expression sites, ranging from complete dissimilarity to the presence of at least some identity. This could permit the trypanosome to keep the M-VSG system distinct from that used in the bloodstream, while allowing a long term drift in the metacyclic repertoire with time mediated by rare conversion by bloodstream VSG genes. We have detected no evidence of ESAGs other than ESAG 1, and believe that although they may be expressed in metacyclic-derived organisms, it is from outwith the M-VSG transcription unit (Graham and Barry, submitted).

Duplicative transposition of the ILTat 1.22 gene

The telomeric nature of the 1.22 basic copy locus, its genomic stability and its ability readily to convert bloodstream VSG gene expression loci in bloodstream infection have implications for the gene conversion process involved in duplicative transposition. It has been suggested previously that the conversion process may have a 3′−5′ orientation with respect to the VSG gene (14), conversion initiating in the telomere repeats or at the chromosome end, as can occur in yeast (41,42). This orientation is not compatible with the restriction of the 1.22 gene to a donor role in duplicative transposition. At its simplest, such a process would result in the telomeric 1.22 basic copy gene being converted as readily as it converts bloodstream VSG gene telomeric loci. As pointed out above, this is not so. Whilst it is formally possible that our approach, the isolation of trypanosomes expressing 1.22, would deny us trypanosomes in which the 1.22 telomere acts as recipient in conversion, the great stability of this telomere argues against this.

Our results also indicate that the extent of homology between VSG gene loci may not be the primary determinant in their frequency of duplicative conversion. The 1.22 gene ES has only 1 complete 70bp repeat unit in common with bloodstream expression sites yet this appears to be used consistently in the fairly frequent generation of a 1.22 ELC. Perhaps then, it is the presence of the 70bp repeat element alone which permits initiation of the conversion events. This has led to the proposal of a specific mechanism for VSG gene conversion (43) arising also from the recent observation that VSG switching occurs naturally at a frequency much higher than background mutational rates (30). In the model, which is based on the double-stranded model for gene conversion in yeast, the 70bp repeat unit is a recognition site for a specific double-strand endonuclease perhaps exposed only to telomeres and present at a low level. The greater the number of repeats, the greater the chance for binding the endonuclease. By the mechanics of the double-strand model, the cut telomere would then be the recipient in gene conversion (44). This would result in both the 1.22 basic copy gene and chromosome internal VSG genes (which also have very limited extents of 70bp repeats (3,18)) being restricted to being donors in gene conversion. As well as keeping the M-VSG system from being homogenized with the bloodstream VSG system, this could bias against the chromosome internalization of telomeric VSG genes. It should be noted that such a scenario does not prevent the involvement of non-70bp repeat sequence in the initiation of conversion (45) nor migration of the conversion branch points where there is extensive similarity between the interacting loci (46). Such events might be responsible for the ability of the ILTat 1.61 genes also to undergo duplicative transposition. Although we have not analysed fully conversion at this locus, in at least one case the conversion process seems to terminate upstream of the 70bp repeats. It is also possible that VSG gene conversion may not involve such a specific mechanism, and may operate rather through preferential targetting of more general recombination processes to the panoply of potential secondary structure that long arrays of 70bp repeats may form.

ACKNOWLEDGEMENTS

We thank J. Kooter and P. Borst for the gift of DNA clones and for sharing unpublished data. We also thank J. Mottram for critical reading of the manuscript. The work was supported by the Medical Research Council and the Wellcome Trust. JDB is a Wellcome Trust senior lecturer.

REFERENCES

1. Vickerman, K. (1985) *British Medical Bulletin*, **41**, 105–114.
2. Pays, E. and Steinert, M. (1988) *Ann. Rev. Genet.*, **22**, 107–126.
3. Van der Ploeg, L.H.T., Valerio, D., De Lange, T., Bernards, A., Borst, P. and Grosveld, F.G. (1982) *Nucleic Acids Res.*, **10**, 5905–5923.
4. Cully, D.F., Ip, H.S. and Cross, G.A.M. (1985) *Cell*, **42**, 173–182.
5. Borst, P. (1986) *Ann. Rev. Biochem.*, **55**, 701–732.
6. Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J. and Borst, P. (1987) *Cell*, **51**, 261–272.
7. Johnson, P.J., Kooter, J.M. and Borst, P. (1987) *Cell*, **51**, 273–281.
8. Alexandre, S., Guyaux, M., Murphy, N.B., Coquelet, H., Pays, A., Steinert, M. and Pays, E. (1988) *Mol. Cell. Biol.*, **8**, 2367–2378.
9. Pays, E., Tebabi, P., Pays, A., Coquelet, H., Revelard, P., Salmon, D. and Steinert, M. (1989) *Cell*, **57**, 835–845.
10. Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N. and Steinert, M. (1985) *Nature*, **316**, 562–564.
11. Bernards, A., DeLange. T., Michels, P.A.M., Lui, A.Y.C., Huisman, M.J. and Borst, P. (1984) *Cell*, **36**, 163–170.
12. Liu, A.Y.C., Van der Ploeg, L.H.T., Rijsewijk, F.A.M. and Borst, P. (1983) *J. Mol. Biol.*, **167**, 57–75.
13. Campbell, D.A., Van Bree, M.P. and Boothroyd, J.C. (1984) *Nucleic Acids Res.*, **12**, 2759–2774.
14. Florent, I., Baltz, T., Raibaud, A. and Eisen, H. (1987) *Gene*, **53**, 55–62.
15. Shah, J.S., Young, J.R., Kimmel, B.E., Jams, K.P. and Williams, R.O. (1987) *Mol. Biochem. Parasitol.*, **24**, 163–174.
16. Aline, R.F. Jnr., MacDonald, G., Brown, E., Allison, J., Myler, P., Rothwell, V. and Stuart, K. (1985) *Nucleic Acids Res.*, **13**, 3161–3177.
17. Michels, P.A.M., Van der Ploeg, L.H.T., Liu, A.Y.C. and Borst P. (1984) *EMBO J.*, **2**, 1345–1351.
18. Aline, R.F., Scholler, K., Nelson, R.G., Agabian, N. and Stuart, K. (1985) *Mol. Biochem. Parasitol.*, **17**, 310–320.
19. Timmers, H., DeLange, T.M., Kooter, J.M. and Borst, P. (1987) *J. Mol.Biol.*, **194**, 81–90.
20. Lenardo, M.J., Rice-Ficht, A.C., Kelly, G., Esser, K.M. and Donelson, J.E. (1984) *Proc. Natl. Acad. Sci.(USA)*, **81**, 6642–6646.
21. Lenardo, M.J., Esser, K.M., Moon, A.M., Van der Ploeg, L.H.T. and Donelson, J.E. (1986) *Mol. Cell. Biol.*, **6**, 1991–1997.
22. Cornelissen, A.W.C.A., Bakkeren, G.A.M., Barry, J.D., Michels, P.A.M. and Borst, P. (1985) *Nucleic Acids Res.*, **13**, 4661–4676.
23. Delauw, M.F., Laurent, M., Paindavoine, P., Aerts, D., Pays, E., Le Ray, D. and Steinert, M. (1987) *Mol. Biochem. Parasitol.*, **23**, 9–17.
24. Barry, J.D. (1986) *Tropical Diseases Bulletin*, **83**, R2-R25.
25. Crowe, J.S., Barry, J.D., Luckins, A.G., Ross, C.A. and Vickerman, K. (1983) *Nature*, **306**, 389–391.
26. Esser, K.M. and Schoenbechler, M.J. (1985) *Science*, **229**, 190–193.
27. Turner, C.M.R., Barry, J.D. and Vickerman, K. (1986) *Parasitology*, **92**, 67–73.
28. Barry, J.D., Crowe, J.S. and Vickerman, K. (1983) *Nature*, **306**, 699–701.
29. Graham, S.V., Matthews, K.R., Shiels, P.G. and Barry, J.D. (1990) *Parasitology*, in press.
30. Turner, C.M.R. and Barry, J.D. (1989) *Parasitology*, **99**, 69–75.
31. Hajduk, S.L. and Vickerman, K. (1981) *Parasitology*, **83**, 609–621.
32. Bernards, A., Van der Ploeg, L.H.T., Liu, A.Y.C., Borst, P., Boothroyd, J.C., Coleman, S. and Cross, G.A.M. (1981) *Cell*, **27**, 497–505.
33. Kaiser, K. and Murray, N.E. (1985) In Glover, D.M. (ed.), DNA Cloning-A Practical Approach. IRL Press, Oxford, Vol.1 pp.1–48.
34. Bernards, A., Kooter, J.M. and Borst, P. (1985) *Mol. Cell. Biol.*, **5**, 545–553.
35. Swinkels, B.W., Gibson, W.C., Osinga, K.A., Kramer, R., Veeneman, G.H., van Boom, J.H. and Borst, P. (1986) *EMBO J.*, **5**, 1291–1298.
36. Feinberg, A.R. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
37. Kimmel,B.E., ole-Moiyoi, O.K. and Young, J.R. (1987) *Mol. Cell. Biol.*, **7**, 1465–1475.
38. Murphy, N.B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M. and Pays, E. (1987) *J. Mol. Biol.*, **195**, 855–871.
39. Barry, J.D., Graham, S.V., Matthews, K.R., Shiels, P.G. and Shonekan, O.A. (1990) *Biochem. Soc. Trans.*, in press.
40. Son, H.J., Cook, G.A., Hall, T. and Donelson, J.E. (1989) *Mol. Biochem. Parasitol.*, **33**, 59–66.
41. Pluta, F. and Zakian, V.A. (1989) *Nature*, **337**, 429–433.
42. Wang, S-S, and Zakian, V.A. (1990) *Nature*, **345**, 456–458.
43. Barry, J.D. (1989) in McAdam, K.P.W.J. (ed), New Strategies in Parasitology. Churchill Livingstone, London, pp 101–116.
44. Fincham, J.R.S. and Oliver,P. (1989) *Nature*, **338**, 14–15.
45. Lee, M.G.S. and Van der Ploeg, L.H.T. (1987) *Mol. Cell Biol.*, **7**. 357–364.
46. Kooter, J.M., Winter, A.J., d'Oliveira, C., Wagter, R. and Borst, P. (1988) *Gene*, **69**, 1–11.